



Not Exactly Rocket Science

« How birds see magnetic fields – an interview with Thorsten Ritz

I've got your missing links right here – 27th November 2010 »

15-minute writing exercise closes the gender gap in university-level physics

By Ed Yong | November 25, 2010 2:00 pm



Think about the things that are important to you. Perhaps you care about creativity, family relationships, your career, or having a sense of humour. Pick two or three of these values and write a few sentences about *why* they are important to you. You have fifteen minutes. It could change your life.

This simple writing exercise may not seem like anything ground-breaking, but its effects speak for themselves. In a university physics class, **Akira Miyake** from the University of Colorado used it to close the gap between male and female performance. In the university's physics course, men typically do better than women but Miyake's study shows that this has nothing to do with innate ability. With nothing but his fifteen-minute exercise, performed twice at the beginning of the year, he virtually abolished the gender divide and allowed the female physicists to challenge their male peers.

The exercise is designed to affirm a person's values, boosting their sense of self-worth and integrity, and reinforcing their belief in themselves. For people who suffer from negative stereotypes, this can make all the difference between success and failure.

Aspiring female scientists and mathematicians still have to contend with the inaccurate stereotype that men are innately better at them in their chosen fields. On top of the challenging nature of their subject, they also have to deal with the dispiriting nature of the stereotype, and the fear that they might live up to it. This problem of "**stereotype threat**" is well known. It catches people in a vicious cycle, where poor

performance leads to greater stress, which leads to poorer performance and even greater stress, and so on. Miyake's exercise is designed to break that cycle.

This isn't the first time it has worked either. It was first used by Geoffrey Cohen (who Miyake works with) to **turn the fortunes of black students in American high schools**. They too face the problem of stereotype threat. In 2007, Cohen showed that his writing drill **boosted the grades of black students**, who still benefited even two years later. The gap between them and their white classmates narrowed and their grade point averages increased, *particularly* among the *weakest* students.

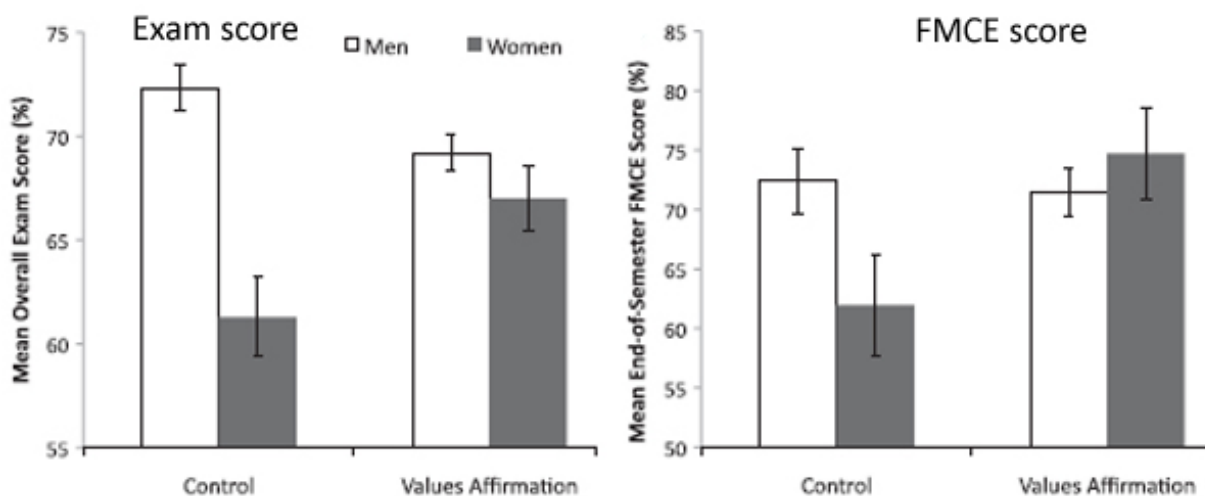
It was a sensational result and the team wanted to see if it could work in other areas. The issue of women in science was an obvious choice. Women still make up a minority of PhD students in physical sciences, maths, engineering and computer sciences. Those that do take up these subjects tend to get lower **grades** during university courses.

To see if their task could help, Miyake recruited 283 men and 116 women who were taking part in the university's 15-week introductory course to physics. He randomly divided them into two groups. One group picked their most important values from a list and wrote about why these mattered to them. The other group – the controls – picked their *least* important values and wrote about why these might matter to *other people*.

This happened twice at the start of the course, and the whole thing was led by teaching assistants who didn't know what was going on (it was a "**double-blind**" experiment). They, and the students, were all told that the exercise was meant to improve writing skills.

The task worked. During the rest of the semester, the students sat for four exams that made up most of their final grade. Among the control group, who wrote about other people's values, men outperformed women by an average of ten percentage points. But among the students who affirmed their own values, the gender gap largely disappeared. Their final grades reflected this shrunken divide: if the women took Miyake's exercise, far more got Bs and far fewer got Cs.

Miyake also gave the students a standard test called the Force and Motion Conceptual Evaluation (FMCE), which checks their understanding of basic physics concepts. In Miyake's control group, the men outscored the women, as they usually do. But the women who wrote about their values closed the gap entirely.



In both cases, the exercise was especially beneficial for women who actually believed that they aren't as good as men at physics. If they bought into the stereotype, even slightly, it cost them dearly in terms of

their scores. Miyake's task provided them with a psychological shield against this threat, allowing them to achieve results on a par with their male classmates.

Like the study with black students, this one shows how pernicious the problem of stereotype threat can be, even among educated, intelligent women who are strongly motivated to learn about their chosen field. It also tells us how easy the threat is to fight.

Miyake's achievement is doubly impressive because the physics course had already tried to introduce ways of reducing the gender gap, including extra tutorials. But all of these methods involved *more* of the same – more teaching, or more problems to solve. Miyake's exercise, by contrast, had nothing whatsoever to do with physics; it worked because it improved *the environment in which women learn physics*. Put it this way: if someone can't hammer in a tricky nail, it might not be because their arm isn't strong enough. It might be that they constantly have to look over their shoulders while they work.

The trick is to intervene at the right time. A scientific education builds on itself, and you need strong foundations to succeed at later levels. In this respect, Miyake thinks that his value-affirming exercise has two benefits: it break the vicious cycle of stereotype threat, but it also sets up a positive cycle too.

Women who are more confident in their own identity do better on the university course, which would boost their confidence further, allowing them to excel at a higher level, and so on. As Miyake says, "Reducing the gender gap at gateways could not only benefit women's performance in the short term but also encourage them to choose and persist in a scientific major and career path in STEM [*Science, Technology, Engineering and Maths*] disciplines."

Reference: Science <http://dx.doi.org/10.1126/science.1195996>

Images: Selected female physicists, clockwise from top-left: *Rosalind Franklin, Sarah Kavassalis, Lise Meitner, Lisa Randall, Caroline Herschel, Reva K. Williams, Maria Mayer, Jocelyn Bell Burnell, Marie Curie and Jennifer Ouellette.*

The original use of the values exercise: [Simple writing exercise helps break vicious cycle that holds back black students](#)

And more on gender equality:

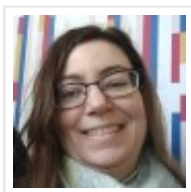
- [Gender gap in maths driven by social factors, not biological differences](#)
- [Women make safer financial decisions when faced with sexual stereotypes](#)
- [Hidden beliefs in science stereotypes predict size of gender gap across 34 countries](#)
- [How objectification silences women – the male glance as a psychological muzzle](#)
- [Mind your words – how stereotypes affect female performance at maths](#)

If the citation link isn't working, [read why here](#)

//



Comments (71)



Sheila

Hiya, offtopic comment. You have a typo in 'boguht'.

Ontopic, Why was the other task related to values and not on some completely unrelated topic?

November 25, 2010 at 2:11 pm

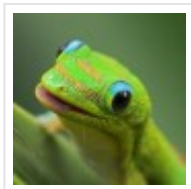


Ed Yong

@Sheila – because if you're designing a control exercise, you want to have something that's as close to the actual one as possible, but leaves out the core element that you're testing. In this case, that element was getting people to affirm their own values and sense of self. The obvious control is to get them to do the exact same thing, but thinking about another person.

If the control was, say, writing about cars, and you saw the same results, you couldn't say if the benefits were due to writing about people, writing about values in general, or any number of possibilities besides writing about personal affirming values.

November 25, 2010 at 2:43 pm



E. coli

Interestingly, it looks like there's a significant difference between the exam scores of the men in the control group and the affirmation group. Did the authors mention this?

November 25, 2010 at 2:54 pm



Ed Yong

@E.coli – Yes they did, but briefly. They say that it's unexpected and that it only applied to the exam grades and not the FMCE scores and the men's grade distribution didn't change.

November 25, 2010 at 3:05 pm



David Dobbs

Ed,

Nicely done. This adds a valuable replication to the Cohen study of African-American highschoolers you mention in your post. Readers wanting more on that might be interested in seeing a 2007 Mind Matters post at SciAm written by Sian Beilock, a U Chicago psychologist who has done some good stereotype threat work herself. The post unfortunately got mangled typographically in one of SciAm's several site overhauls the last few years, but is interesting nonetheless.

<http://www.scientificamerican.com/blog/post.cfm?id=closing-the-racial-achievement-gap>

November 25, 2010 at 3:28 pm



Elaine

This is an amazing study. Valuing values – allowing time for the students to consider what is important and why would I guess be a regular reminder about who they are and what they want to achieve.

November 25, 2010 at 3:50 pm



QoB

This is so great to see.

November 25, 2010 at 6:15 pm



WhatMeWorry

So how about some 15 minute tests that can be taken by men, before exams on which women typically score better (just about everything else), to boost the men's scores? As I understand it, the biggest problems facing university these days is NOT underperforming women in physics, it's about underperforming men in just about every other discipline.

And did you consider what might happen to the men's scores on the physics tests had they too taken a confidence-building test?

And now that you have 'proved' women can score as highly as men on physics exams, and they score higher on just about everything else, that just makes them flat-out superior homo sapiens, right?

November 25, 2010 at 8:16 pm



somecomment

The same way SAT exams have been dumbed down to close the gender gap, expect these exercises to be compulsory in American schools. Of course, nothing will be done about topics where women outperform men, such as the ones related with language. And no research will be devoted to find exercises for men to improve their grades in physics.

When women are better, it is because they are better. When men are better, it is because society oppresses women. Conclusion: women are better. This is called "equality".

Lots of studies show that men are better in STEM than women. None of these studies will be published in Discover Magazine or other politically correct sites on the Internet. But the moment ONE study goes the opposite direction, it is

published immediately. This is called “neutrality” and science is known because of it.

Meanwhile, no matter what measures they are trying to apply, men would keep on outperforming women in physics, forty years from now. Want to bet?

November 25, 2010 at 8:50 pm



Coturnix

Where are all these commenters suffering from insecure masculinity coming from? A cave?

November 25, 2010 at 9:55 pm



Lyr

Thanks for mansplaining that to us, somecomment. 😊

November 25, 2010 at 10:03 pm



Scott

WhatMeWorry, read the article more carefully.

“Miyake recruited 283 men and 116 women who were taking part in the university’s 15-week introductory course to physics. He randomly divided them into two groups. One group picked their most important values from a list and wrote about why these mattered to them. The other group – the controls – picked their least important values and wrote about why these might matter to other people.”

The men also took the confidence building test.

somecomment, I will take up your bet. I will also bet that the gap is DECREASED in forty years, though, and that it will continue decreasing after that.

November 25, 2010 at 10:58 pm



Scott

Also, WhatMeWorry, no one is saying that women are the superior homo sapiens. Did you know that other studies have shown that in gender-equal situations men are just as empathetic as women, and just as verbally fluent? Even though in typical ones they test worse. Stereotypes really do work both ways, and they don’t make anybody BETTER, just half the population worse in every area they affect. That’s the problem. It’s not a gender war thing, the genders should work together so that men can be better at communicating and women at calculating.

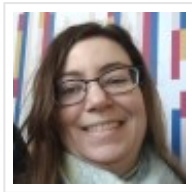
November 25, 2010 at 11:02 pm



Kim H

Huh. It sounds like an exercise that should be done constantly, if it's especially effective for students who believe for some reason that they aren't good at (whatever the activity is). I wonder whether there are any cases in which focusing on students' values would make it more difficult to learn science – if there's a clear benefit to students who think they can't do it (and that might include lots of other groups besides women), and no harm to students who are already confident, then it would be worth doing during the first week of class.

November 25, 2010 at 11:07 pm



Sheila

@Ed, I was wondering how they would be able to tell that the result comes from focusing on positive personal value rather than writing anything or even some other distractor task.

Oh, and if it's from focusing on the topic, I wonder if it would work in other modalities, like a group discussion beforehand.

It's interesting stuff! I like pondering other experiments they could try.

November 26, 2010 at 4:29 am



Neil

"[...]students were all told that the exercise was meant to improve writing skills."

Will it still work if the students are told the reason behind the exercise?

Doing experiments on/with humans must be no fun at all – give me 10^{20} numbers of non-sentient things any time!

November 26, 2010 at 5:32 am



Ed Yong

@Scott said, "Read the article more carefully... The men also took the confidence building test." Exactly. This is the sort of insight you can get from actually *reading the post* and not just beating the keyboard with your fists and face.

@KimH – I suspect that doing it many times over would make students suspicious about it, but there could be other ways of achieving the same result without doing the exact same task.

@Sheila – If the effect was just down to writing or doing something distracting, the control group would have experienced benefits too, because that's what they did.

@Neil – Good question. I suspect not. Doing that would just bring all the issues of stereotype threat to the fore. It would be a bit like saying, "I'm now going to show you how to solve this really, really difficult problem. Don't panic." Although obviously that's something that's worth investigating.

November 26, 2010 at 5:43 am



Ian Eiloart

I guess the conclusion to draw is that men benefit from writing about the values of others, while women benefit even more from writing about their own values. Although, because all of the subjects wrote about values, we can't assume that the exercises actually helped anyone.

These results also support the inverse hypothesis: that writing about the values of others harms women, and writing about their own values harms men.

It would be interesting to see the outcomes when men and women undertake both exercises, and when they undertake neither exercise (perhaps writing instead about attributes of an experiment, or a transport system, or something else non-human).

If we do assume that these exercises produce a better outcome than no exercise, then the lesson must be to ask men to write about values of others, and women to write about their own values. Or to undertake both exercises.

Why would men benefit by writing about the values of others? Perhaps it helps them to empathise with others better, and therefore helps them with collaborative learning elements of the later course.

November 26, 2010 at 7:02 am

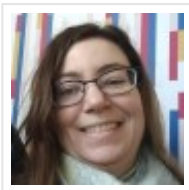


Rich

@4. Ed Yong: "the men's grade distribution didn't change."

How do you get a fall in the mean of 5%-6% without the distribution changing?

November 26, 2010 at 8:49 am



Sheila

Oh, I see your point. doh.

November 26, 2010 at 10:12 am



Bjørn Østman

@4. Echoing Rich, I also wonder how one can say the men's grade distribution didn't change. It looks like a significant decrease in the men's exam scores. Looks like they shouldn't be writing about their values.

November 26, 2010 at 1:31 pm



Huh

“But among the students who affirmed their own values, the gender gap largely disappeared.”

No it didn't, the gender gap is still there, it just reversed and the vast majority of students lost ground. What is their hypothesis? Duh, we don't know is not good enough.

Methinks Hans the Clever Horse could figure this one out. Can you, Miyake? Can you, class?

November 26, 2010 at 2:11 pm



Lyr

Women are generally socialized to think of others first, while men are not. This could be why the results of the exercise were different for the two genders.

November 26, 2010 at 6:26 pm



Nullius in Verba

How did they calculate the error bars? What do they represent? 95%?

The error bars seem to be about (+/-) 1-2% in size. Since the sample sizes were presumably around 58 and 141 (although intriguingly they seem to be *different* for the control and value-affirming groups), the spread is reduced by a factor of about 7 or 12 by the averaging process. That suggests there is quite a broad spread in the raw data.

Presumably this isn't the uncertainty in the exam score as a measurement of performance (or they'd be useless as exams). So my guess would be that there is a spread of abilities, from high to low, and the concern being considered here is that more than the expected number of low ability students coincidentally get picked for the control group.

There are lots of ways that odd distributions of ability could mess things up. For example, if most students score about 70 with only a few very weak students who would score very low, the odds of most of those ending up in the control group by chance are significant. You could estimate the spread of abilities by looking at the exam results.

Assuming the statistics are calculated correctly, there does appear to be an effect here. But the rest of the discussion – that the difference is due to stereotyping, that the difference is due to women thinking they can't do physics, that the effect works by boosting self-confidence – all seem to be unsupported speculation. Is there any evidence for these statements?

November 26, 2010 at 7:04 pm



James

“The same way SAT exams have been dumbed down to close the gender gap, expect these exercises to be compulsory in American schools.”

Um, the SAT hasn't been dumbed down. It's substantially more difficult than it was 20 years ago, with a large essay portion, higher-level math, and elimination of the (trivial) analogies section.

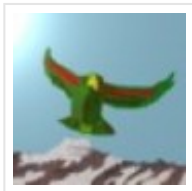
November 26, 2010 at 8:45 pm



Ed Yong

@James – Spoilsport. It's hardly fair to expect Somecomment to let pesky facts get in the way of his ranting...

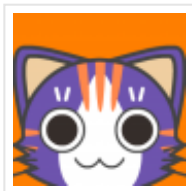
November 26, 2010 at 8:57 pm



Kea

You could have tried harder to use photographs of professional physicists.

November 26, 2010 at 9:06 pm



katt

@Nullius in Verba

“Assuming the statistics are calculated correctly, there does appear to be an effect here. But the rest of the discussion – that the difference is due to stereotyping, that the difference is due to women thinking they can't do physics, that the effect works by boosting self-confidence – all seem to be unsupported speculation. Is there any evidence for these statements?”

I can't speak for the statistics, but the phenomenon of stereotype threat is very very well documented, especially with regards to women and performance on math/logic/scientific-type tasks. It is eminently plausible, given the literature and the design of this experiment (similar to the one used with black students, another population in which the effects of stereotype threat are well documented), that the female physics students' improvement was due to mitigation of stereotype threat by boosting their feelings of self-efficacy in some way. The burden's on you to think of a better explanation for the results here.

And while there is a debate on how much of gender achievement gaps stereotype threat explains and the exact mechanisms by which it works, especially in more complex non-academic situations, understanding this debate properly requires literacy in the basic background stereotype threat research. It's clear to me, by calling stereotype threat-related explanations “unsupported speculation”, that you are completely unfamiliar with the topic in general. I'd suggest you punch “stereotype threat” into google and curl up with some eggnog and start reading.

Here's a good place to start:

<http://www.reducingstereotypethreat.org/definition.html>

November 27, 2010 at 1:04 am



Abu Dhabi

What I'd like to know is how the gender distribution looked among the two groups (said to have been divided up randomly), and how the individual grades of the students /changed/. Without that information, I'm forced to conclude it plausible that the differences arise merely from random chance, not any effect of the exercise itself. I'd go look at the paper itself, but I'm not shelling out fifteen bucks for one paper.

November 27, 2010 at 4:42 am



Eleanor

I'm pretty gobsmacked at the level of change – I've been aware of this kind of study, but, wow. Reading the links to other studies, I shouldn't have been that surprised.

Of course, as pointed out by some of the other commentators, this is also a commie plot to turn women into men and men into vegetarian, Guardian-reading liberals, thus distroying the fabric of Western society as we know it, so Down With This Sort of Thing.

November 27, 2010 at 6:24 am



Nullius in Verba

#28,

I don't doubt that stereotype threat is very well documented. But given the amount of work done on it, one would therefore expect there to be well-developed methods for identifying it precisely and distinguishing it from other possible causes.

Writing an essay talking about values important to oneself could have a variety of psychological effects: long-term versus short-term planning, seriousness versus playfulness, industriousness versus laziness, calm versus excitement, caution versus boldness, defiance versus surrender, analytical observation versus imagination and projection, an inwards versus an outwards focus, the activation of moral principles and values that encourage hard work and being careful. Have we tested that the "control" intervention of writing an essay about other people's values didn't have a *negative* effect on the women? What values did people write about, that were being activated or reinforced? I have no idea whether any of these are "better" explanations, but I don't see any reasons given here to suppose they are worse, either.

The argument that "we can't think of any other explanation therefore the hypothesis must be true" is a fallacy known as *argument from ignorance*. The argument that "seeing the expected consequences of a hypothesis is evidence for the hypothesis" is a fallacy known as *confirming the consequent*. An explanation or proposed mechanism being "eminently plausible" isn't sufficient.

I recall reading of another experiment where they gave people two tests. After the first, they told half the subjects at random that they had scored highly and well above average (irrespective of whether they actually had), and the other half that they had scored very low and their result was inferior. Members of the first group scored higher on the second test than they did on the first, and members of the second scored lower. That also has other possible explanations (such as resentment against an evident injustice or whether you take the second test seriously), but the cause is more directly connected to self-confidence and performance expectations. Humans are complicated. You have to do a lot of work to make sure you're testing what you think you're testing.

I'd suggest reading Feynman's essay on **cargo cult science** (the suggestion is not intended in any sort of adversarial sense) and in particular the discussion of psychological experiments on rats running mazes. It's a good explanation of what I'm talking about.

Note, I'm not saying in any of this that the experimenters here *haven't* considered all this, that they don't have other experiments to back up the connection, or that this is an example of cargo cult science. I was just asking for more detail.

November 27, 2010 at 6:40 am



My Posting Career

It's pretty sad that there are so many people willing to suspend their critical faculties, but this is the end result of years of PC dominance in academia. Of course it defies common sense to say that doing some essay on values makes women's brains work better, and the people here cheering this seem wilfully unaware of the long history of basically fraudulent social science on this subject. So many basic problems with this study that it mainly demonstrates how powerful the will to believe is.

Yes, children, blacks are just as smart as other races, they just were never encouraged enough (see glorious sub-Saharan Africa). Women are just as capable as men at math, but being forced to play with dolls makes them dumb. Etc, ad nauseam. When it comes down to it, people are made dumb by ideology.

November 27, 2010 at 11:12 am



Lyr

My Posting Career doesn't sound like he's too fond of blacks or women, or the thought that they might be equal in intelligence to him. He wouldn't have felt the need to bring up what he did in the second paragraph of his post otherwise. And now he is blaming political correctness for why women might have seemed to do better after the writing exercise...but if the results discussed in the original article are all due to political correctness in academia, how do you explain all the data backing up stereotype threat? The answer is that you can't.

And before he says "well if women are as good at science as men, where are all the women scientists", I'll answer that with the old punchline "they're with all the

black Nascar drivers". Culture plays a big role in how we view our abilities and what careers we choose.

November 27, 2010 at 11:44 am



Question

How did they get all the students to actually participate? I find it hard to believe that all of the students would actually participate.

The description of the study as double blind seems to be a bit of a stretch...

p.s I wish my physics courses had a B average.

November 27, 2010 at 12:08 pm



Nullius in Verba

#32, 33,

There are more than just two ways to interpret the observed results. You can't assume that if some subset of the population don't get such good results, that it is because they are innately less intelligent. Nor can you assume that it is because they are being discriminated against through stereotyping. Saying that you doubt explanation A doesn't imply you support explanation B. Disproving explanation B does not thereby prove explanation A. We have explanations C, D, E, F..., and so on that you're not even considering.

The first thing you have to do when considering such a question is to be genuinely open-minded to the possibility that the theory you really don't like might be true. For some people, this may be that women really could be inherently less intelligent. There's no law of nature that says they have to be the same – just as women tend to be shorter or have better colour vision. For others, it may be that the differences really are to do with irrational discrimination. Everybody has cognitive biases; it's the way humans are built.

And one of those biases is confirmation bias – we examine theories we don't like or that conflict with our current beliefs more critically than those that we are already inclined to believe. In science, we must guard carefully against it, and the first and most important step to doing so is to recognise and acknowledge that we have it. More, that all the other scientists have it, and that despite all our efforts to keep it out it inevitably pervades even published and widely-accepted results.

All the evidence I am aware of says that men and women have indistinguishable innate ability to do maths and physics – all the reasons for the differences in outcome are social and psychological. The same goes for a lot of other characteristics and skills that have been linked with sex: capacity for violence, emotional sensitivity, criminality, nurturing, multi-tasking, etc. The fact that the vast majority of people sent to prison are male is not biologically built-in. The fact that the vast majority of scientists are male is not biologically built-in.

That does *not* imply, however, that the reasons are to do with sexist discrimination, or stereotyped gender roles imposed on people by the pressure of

social expectations. Or at least, not directly. People make different choices, have different values, get different opportunities, and there are many confounding factors. People are often different because they *want* to be, because it is to their advantage. People are different due to accidents of history, due to the division of labour, due to separate social groups being partially isolated from one another and therefore taking separate random walks through custom and fashion.

Nor should you assume that such differences are necessarily to a particular group's disadvantage, or that it would be better if we were all the same. Such value judgements are in any case outside the remit of science, and increase the risk of confirmation bias.

To explain all the data backing up stereotype threat, first we would have to know what it is. That was what I was asking. I'm sufficiently convinced that the cause is social/psychological, but I have not seen evidence that it is stereotype threat, and at first glance it does look like an attempt to rescue a hypothesis in circumstances where it doesn't appear to apply. Particularly when you see that most of the research has concentrated on the politically popular areas of race, sex, and gender, and not on the multitude of other groupings and social expectations that one might expect to show similar effects. (e.g. introvert/extrovert, physical/logical/artistic, young/old, attractive/unattractive, follower/leader, conservative/radical, authoritarian/libertarian, and so on.)

But you have presumably already considered this, and can briefly summarise the evidence and logic that led you to the conclusion that it was stereotype threat in the case above. I didn't think it was obvious, and others are clearly dubious too, but maybe there's a simple explanation?

November 27, 2010 at 2:14 pm

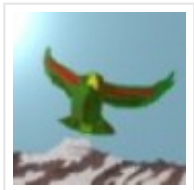


Bram

I'm no statistics major, but I suppose Nullius in Verba is raising some good questions?

Perhaps Myiake et al. are reading NERS? Wouldn't be the first time authors of a study would come over comment. Seems to me they're people to answer said questions.

November 27, 2010 at 3:19 pm



Kea

Eleanor, Western society as you know it has already been destroyed. Wake up.

November 27, 2010 at 3:29 pm

From FMCE score and Exam score , on average the control group women scored almost 10% below men and the value affirmation group scored equal. It is a very



Valli

interesting study and I am encouraged by this finding. But the accuracy of the results depend on how 'randomness' is assured in picking the members of the two groups.

Self esteem plays an important role in performance. Normally people with same values group together and sit together . That would group men with superior attitudes with women with problems in self esteem and would group well adjusted men and women together. If the 2 sets of papers are distributed in the class to two different sections , the resulting groups may not be truly random. Though I am writing here a possible scenario , I like the researchers to look in closely at possible non-randomness to their group makeup.

It is hard to believe a 15mins write up can effect a 10% grading gap change in college entry Physics students who have gone through years of education and competition to get there. Schools here do focus to boost the morale of its students by many means including essay writing.

November 27, 2010 at 8:26 pm



katt

@ Nullius:

You're hitting on about a bazillion different criticisms of this paper here. I'll try to address the ones I can pick out.

"well-developed methods for identifying it", "first we would have to know what it is": There aren't really any well-developed methods for identifying *anything* in psychology. Science in general is a messy practice and psychology may be the messiest practice of them all. Whether that makes psychology a "cargo-cult science", I don't know. As someone who studies across several disciplines I frequently get angry at poorly constructed experiments, the general tendency in psychology not to publish replications, badly defined experimental constructs, lack of mechanism or theory behind an effect, dressed-up exploratory studies, etc. But these are general criticisms of the field, and not this particular experiment, so I'll leave them behind for now.

Since stereotype threat is a construct that operates across several different levels of explanation (large-scale social norms, small-scale social perception and interaction, individual interaction with the environment, intra-individual attitudes and cognitive function) it is particularly hard to "pin down" as it were as it isn't a "single thing" that you can immediately perceive upon looking at a situation or set of data; instead, it operates at the nexus of a large body of psychological forces. You can examine whether stereotype threat is present in an environment by looking for markers of these forces, i.e. to see if group-identification is made salient, that a stereotype is associated with that group, that individuals with that group identification feel pressure or anxiety about potentially conforming to that stereotype, that individuals do in fact perform negatively when all of this is in place, etc. but you can't necessarily quantify it very well (even if you do make a scale or something it will always be based upon the former sorts of assessments).

So to run through what they did check:

They didn't measure this, but I think it is plausible that group-identification is made salient in physics classes (there are visibly fewer women than men in these classes). It is well-known that this stereotype of women exists in the general population, and the students were asked the following question on a 5-point scale: "according to my own personal beliefs, I expect men to generally do better in physics than women". Some endorsed it, showing that this stereotype is known and affirmed amongst women taking physics. Most importantly for the stereotype-threat hypothesis, this is what the researchers found: as women endorsed the stereotypical view more, their exam scores decreased in the control condition; there was no such relationship between endorsing the stereotype and performance in the affirmation condition; the improvement effect happens in the group of women who highly endorse the stereotype; and endorsement of the stereotype has no effect on men's performance whatsoever.

So we have 1. markers that would suggest that stereotype or identity threat is present, 2. the performance of female but not male students varying with the presence of markers of stereotype threat and 3. the experimental intervention erasing the effect seen in 2, the intervention being more effective amongst those that have the markers from 1.

I think this constitutes decent evidence for the operation of stereotype threat and for the subsequent mitigation of its effects. How about you?

Since this post is long I am going to comment again about mechanisms of stereotype threat

November 28, 2010 at 4:24 am



katt

"Writing an essay talking about values important to oneself could have a variety of psychological effects: long-term versus short-term planning, seriousness versus playfulness, industriousness versus laziness, calm versus excitement, caution versus boldness, defiance versus surrender, analytical observation versus imagination and projection, an inwards versus an outwards focus, the activation of moral principles and values that encourage hard work and being careful."

Stereotype threat is a complicated construct and admittedly, no one quite knows the specifics of how it operates within an individual (i.e. what cognitive, affective, neurobiological etc. effects it has such that it produces poor performance on certain tasks, negative life outcomes, disproportionate representation in academic fields and in different professions and so forth). Many of the explanations you propose are completely compatible with a stereotype-threat explanation: for example writing about one's own values might indeed induce a sort of self-focused meditative state compared to writing about other people's values, and this might mitigate the performance anxiety caused by a stereotype being particularly salient. I'm going to refer you to that site I did before; they have an excellent section on work done on mechanisms behind stereotype threat:

<http://www.reducingstereotypethreat.org/mechanisms.html>

"That does not imply, however, that the reasons are to do with sexist discrimination, or stereotyped gender roles imposed on people by the pressure of

social expectations. Or at least, not directly. People make different choices, have different values, get different opportunities, and there are many confounding factors. People are often different because they want to be, because it is to their advantage. People are different due to accidents of history, due to the division of labour, due to separate social groups being partially isolated from one another and therefore taking separate random walks through custom and fashion.”

I’m sure you know that we make our choices within the context of societal forces. What is to our advantage, what happened because of accidents, etc. is the background against which we make our choices. A woman interested in sciency-stuff will find it much easier to go into the biological sciences than she will physics; in the humanities, into literary criticism than philosophy. There are far more women in cognitive science than in philosophy of mind, for example, despite both fields being different approaches to roughly the same problems. The effects of sexism aren’t always direct (as yourself said); women and other people affected by the -isms don’t want to fight every little thing and second-guess themselves about every choice they make (am I quitting this job because I hate it, full stop, or am I quitting because I hate it because my boss makes me run register instead of the men?). However implicit these forces may be in our decisions and however impractical or practical it may be for us to individually examine them and their causes, it is possible though to examine them on the whole through experimentation and to tease out their causes. What I have seen on research on gender discrimination, as well as in my personal experience, convinces me that there is a pattern of discrimination against women that systematically disadvantages them compared to men, i.e. sexism exists. Given the research I’m familiar with with regards to math performance, it is plausible to me that the same effects are in operation in physics classes and that we have a classic example here of the broader patterns of sexism (the existence and perpetuation of these stereotypes) causing local effects (gender gaps in introductory physics performance).

“particularly when you see that most of the research has concentrated on the politically popular areas of race, sex, and gender, and not on the multitude of other groupings and social expectations that one might expect to show similar effects. (e.g. introvert/extrovert, physical/logical/artistic, young/old, attractive/unattractive, follower/leader, conservative/radical, authoritarian/libertarian, and so on.)”

Stereotype threat has been demonstrated with regards to older people. I’m not sure about the other groups, but stereotype threat has to be a *threat*, it has to cause anxiety about identification with the group and about potentially confirming a negative stereotype associated with it. Thus you wouldn’t expect to see stereotype or identity threat when there are no negative stakes associated with identifying with the group, when identification with the group is low, and so forth. Race and sex/gender are constantly associated with stereotype threat research because 1. group identification is typically highly visible and is made salient in daily interaction 2. pervasive and well-known detrimental stereotypes exist about these groups 3. preventing or mitigating the effects of stereotype threat amongst these groups is a high priority for many researchers. However, making group identification for individuals in any sort of group (even randomly assigned ones) salient can and does have interesting effects. Introductory materials for social

psychology should cover some of these experiments if you aren't familiar with them already.

And finally:

"There are more than just two ways to interpret the observed results. You can't assume that if some subset of the population don't get such good results, that it is because they are innately less intelligent. Nor can you assume that it is because they are being discriminated against through stereotyping. Saying that you doubt explanation A doesn't imply you support explanation B. Disproving explanation B does not thereby prove explanation A. We have explanations C, D, E, F..., and so on that you're not even considering."

The corollary to being critical about explanations and attempting to avoid confirmation bias is making sure that one is equitably critical of each and every study one comes across. It may be the case that this study, or any particular study on sexism or stereotype threat was not well conducted, but if you have a pattern of deeply criticizing these sorts of studies and not studies on, let's say, theory of mind performance in preschoolers, then that speaks of bias, even if those sexism experiments were truly flawed. I'm afraid you might want to consider your biases because you seem to be repeatedly suggesting that there is no basis or no logic behind the idea of stereotype threat in this situation, when it the logic is on its face plausible (it may not ultimately be right, but it certainly is not unjustified).

I gave you a link, Google Scholar is seconds away; it is now time for you to figure out for yourself how plausible this whole business is and to come up with substantive criticisms. This requires at least identifying what is precisely wrong with this particular experiment, the explanation of its results, or the construct of stereotype threat in general. Even better would be offering an alternate explanation and an empirical test to distinguish between your explanation and theirs. This is the way science works. Though thoughtfulness and care in designing experiments and theories are needed, we can't always sit on our butts and think of every possible explanation, variable, or confound before we run an experiment. That's why we run them and then *afterward* people pick them apart, and that's why there are people who do mostly or entirely theoretical work in specific science fields. Science is collaborative, not just in the sense of a group or team of researchers in a lab working on a project but in the sense of these teams and individuals responding to each other and weaving webs of criticism, counter-criticism, alternative theories, etc. One cannot expect the authors of any study to have done it all; instead, it is the job of science as a whole to conduct theory-formation and testing.

I hope I've answered some of your questions, and I hope more strongly that you aren't a very articulate troll looking to provoke people about sexism because I just wrote a lot of words and somebody better read them.

November 28, 2010 at 5:47 am

#38,



Nullius in Verba

Nowadays, most researchers assure randomness by getting the computer to do it. People can still go wrong, but it's so easy to do it right that I'd be surprised if that was a problem.

The question, though, is whether the randomness is able to give the two groups the same average (to within the known margin of error) when all conditions are the same. The other question is whether all the conditions besides the one under test were the same for the two groups.

Suppose for a moment that the error margins might have been calculated wrongly, and that on the Exam the men would score 70 ± 2 and the women 65 ± 4 . Looking at the above figures, you can see that the difference could arise purely by chance. On the FMCE test, the margins would have to be wider – 70 ± 10 , say. If those error bounds drawn on the bar chart were a factor of two out (or if they were supposed to be 1 standard deviation instead of 95%, as is more commonly used), the significance of the result would disappear. It would be random noise, that just coincidentally showed a difference.

Seriously, I think it more likely than not that they're correct, but interpretation does depend on exactly what they are and how they're calculated. There are several different reasonable ways it could be done (corresponding to different null hypotheses).

The other question is whether the two groups were treated differently in any other sense. If the two groups were not only given different essays at the start of the course, but were also taught together as groups, then some extraneous factor (like a particular teacher) could have shifted the scores of one group. Again, it's not too hard to avoid that by good experimental design, but practical considerations often intrude.

The result is odd for another reason. In order to get on the course, the students have all had to pass multiple exams *already*. The women wouldn't have got on the course – probably wouldn't have chosen to do physics – if they hadn't already shown that they could pass physics exams. And unless the university is using 'affirmative action', they must have scored comparably to the men. The evidence that such a selection has taken place is shown by the different population sizes – 283 versus 116.

So you're not really testing 'men' and 'women', you're testing men and women who have elected to study physics and have passed the university entrance qualifications, which is different. If you take two identical bell-shaped distributions, and scale one of them down (to reflect the smaller number of women who *choose* to do physics), then the distribution narrows – you get fewer extremes. (Because the best of the women are more likely to be studying something else.) Now if you cut off the lower half of each bell to reflect the entrance conditions, you find that the mean of what is left is higher for the men simply because there are more of them, and therefore more of them at the extreme high end.

Thus, the difference observed may be nothing at all to do with the different abilities or self-confidence of men and women, but simply because there are more

men who wanted and were able to get on the course. Even without any bias or lack of confidence on the part of women, a gender gap is expected, so what does it mean for it to disappear?

So you see, the brief summary above reports a gender gap in the control group, but ascribes it without discussion to this ‘stereotype threat’ hypothesis. It then applies an intervention consisting of a couple of 15 minute sessions at the start of the course – no discussion of how it works, how long the effect lasts, how we know – and finds the gender gap vanishes. Not only that, but it appears to have a significant *adverse* effect on the men, even though you would imagine some of the men would suffer from stereotype threat based on different stereotypes (e.g. rich/poor background). That does suggest that the error bounds might have been underestimated. We’re not told why this intervention would target stereotype threat in particular, rather than some other psychological factor. It is just asserted, as if this is obvious. And finally, the conclusion supports a popular political belief system so there is an increased danger of confirmation bias.

It’s an interesting result, certainly. And I wouldn’t expect all the details needed for the technical paper to appear in a blog post or popular media article. But there is reason for scepticism, too, and one ought to ask a *few* questions before accepting it as good science.

#40, 41,

Thanks for the comments. I hadn’t seen it when I hit ‘send’ on the above. I’ll reply more fully in a bit.

November 28, 2010 at 5:48 am



Nullius in Verba

katt,

#40,

I agree with your point on the general characteristics of the field. Feynman thought it was possible for psychology to be done right – he speaks of Young’s rat-running experiment as “A-number-one” – and it is quite clearly far harder than physics. But I’d say that general doubts about the field as a whole should be applied to all results from the field, not none.

I agree that any psychological effect will be hard to pin down, and will interact with many other effects. In this case, I thought the effect was sufficiently well-defined. It’s a sort of nocebo effect: low expectations lead to a reduction in performance. Identifying that it is present is a necessary step, but showing that it (or the nexus of factors closely associated with it) is the only effect at work is more important.

Thanks for the summary of the checks they did. The statement “I expect men to generally do better in physics than women” is not really a stereotype, but a statistical observation. We’re all commenting on and criticising the existence of the gender gap here – does this mean we are stereotyping women? The stereotype would be to say that this difference is innate. Nor does it imply that the student

feels threatened by it, or feels that it applies to them. If you asked me whether I endorsed the statement “I expect men to be sent to prison more often than women” I would agree, but I wouldn’t feel any more inclined to criminality as a result. (I don’t think I would, anyway. It would be an interesting experiment.) It applies to a different subset of men; a group that I am not a member of.

The correlation between endorsing the stereotype and exam performance is more interesting. That does suggest a connection, although the causal arrow could go either way. Poor performance leads people to seek out excuses, to bolster their self-image.

That the men don’t show the same relationship only shows they don’t suffer from sex-based stereotype threat. (Or are they are not inclined to use it as an excuse.) There are lots of other stereotypes that they might feel threatened by. Not much to be deduced there.

So, there is some ambiguous support for statement 1., I agree with statement 2. although I caution against correlation implying causation, but the best evidence comes from the second half of statement 3.: that the intervention affected the women with the strongest sex-based performance expectations the most.

I agree that there is some evidence for it, although it is ambiguous and I’d say better controls against possible excuse-making behaviour are needed. To test for that, I might look at the women at the top of the class who endorsed the statement – if they also improved in performance more, then that would be good evidence.

November 28, 2010 at 7:51 am



Nullius in Verba

#41,

The following discussion is less scientific and more political.

“What I have seen on research on gender discrimination, as well as in my personal experience, convinces me that there is a pattern of discrimination against women that systematically disadvantages them compared to men, i.e. sexism exists.”

Yes, there is. But society is complex, and in other areas there is a pattern of discrimination against men that systematically disadvantages them compared to women. Both occur, and which you experience depends on where you are and what you’re doing.

I don’t doubt that sexist discrimination exists, but it isn’t the only factor that can lead to differences of outcome. From my own personal experience, I’ve met plenty of women mathematicians and I haven’t noticed any lack of confidence on their part, or overt sexism on the part of their colleagues. Differences do exist, but there are other reasons evident. For example: there are fewer women in senior positions, but this seems mainly to be due to their unwillingness to play office politics or take on all the bureaucracy – a characteristic shared by a lot of men who also do not ‘get on’.

That's not necessarily a bad thing, either. Those senior positions are highly pressured and involve a lot of unpleasantness and stress, and I purposely steer well clear of it. I can certainly understand women wanting to steer clear too, if they have the financial freedom to do so. So given that my motivation is (hypothetically) exactly the same as the women, and this is often cited elsewhere as an example of sexual discrimination, am I being discriminated against too?

It may be worth noting that a lot of men are actually quite pleased that women have entered the workforce, and have forced employers to adapt, because it means that working conditions that men have had to put up with for decades have been relaxed. Men here can claim paid paternity leave, for example, which would have been unthinkable before equality law and feminism. A lot of bullying that used to happen no longer does.

"I'm afraid you might want to consider your biases because you seem to be repeatedly suggesting that there is no basis or no logic behind the idea of stereotype threat in this situation, when it the logic is on its face plausible (it may not ultimately be right, but it certainly is not unjustified)."

I do try to consider my biases, and I acknowledge that I have them. That's why I talk to people who hold different views, and ask for their reasons.

I don't know whether you would class me as a troll for it, but my aim in raising objections is to get people to be more sceptical (which is not the same thing as disbelieving) about scientific claims in general. The idea is to illustrate the sort of questions we ought to ask, the sort of possibilities we need to consider. As Feynman put it: *"It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty – a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid – not only what you think is right about it: other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked – to make sure the other fellow can tell they have been eliminated."* I will grant you that I'm more likely to do it for some subjects than others, but I'm sure I can rely on other people to help me fill in the gaps in my own confirmation bias, as I seek to help them. 😊

Anyway, it's been a good discussion. My thanks.

November 28, 2010 at 8:03 am



Athene Donald

I am surprised none of the responses have mentioned the book *Whistling Vivaldi* by Claude Steele – who was the first psychologist to identify stereotype threat. The self-affirmation exercise you describe in the context of women in physics is something that has been shown to be effective in a variety of threat situations. Steele describes many examples of studies demonstrating the strength of stereotype threat – he was particularly concerned with black students in US universities, but it also applies apparently to white athletes competing in sprint races against blacks (where the whites are 'expected' to do worse) or the elderly on memory tests. I recommend people to read the book, which is immensely

readable. I recently wrote a blog post about stereotype threat more generally in the context of how it may affect women in science

<http://athenedonald.wordpress.com/2010/11/18/stereotype-threat-underperformance-and-diversity/>, but I still find it surprising how such a ‘small’ thing as a 15 minute exercise can apparently have such a large effect.

November 28, 2010 at 11:18 am



becca

I am inclined to think that the control group is a poor choice.

Last week, in my toastmasters (public speaking) group, for a tabletops (impromptu speaking practice) we had to write down our responses to a list of ‘either/or’ or ‘for/against’ topics (e.g. do you prefer cats or dogs? are you for or against the new TSA screening?). We wrote down our answers based on what we felt... Then, the person leading the exercise turned it around on us. We had to argue *against* the position we had just taken.

Many people found the exercise *very* challenging. (I, loving to play devils advocate the way I do, only found the questions on which I had *no* opinion to be challenging, but I might have been the only one that felt that way). In other words, I think that in this experiment what you’ve got here is two treatment groups: one in which people are given an easier, self-affirming writing assignment, and the other in which people are getting a challenging, self-neutral (or, arguably, even self-refuting) assignment.

In other words, you’re *prompting* the control groups to feel challenged- perhaps even frustrated. Maybe not a good mindset to start off a physics course with.

Of course, most of my toastmasters group pointed out how challenging the exercise was... but didn’t seem upset by it. These folks like a good challenge. It’s possible that doing a non-graded (I presume) exercise in a classroom setting that provided a mental challenge would make most people better equipped to do some physics. This would be nice, because it would mean I am not a weirdo and the results of the study (that my gut reaction bias is to believe) are valid.

@Nullius in Verba

You read: “according to my own personal beliefs, I expect men to generally do better in physics than women” and quoted it as ““I expect men to generally do better in physics than women”. So you (intentionally?) left out psychologically-salient qualifiers.

... You keep on using that word (“cognitive bias”). I do not think it means what you think it means.

November 28, 2010 at 11:42 am



Nullius in Verba

#45,

How does “according to my own personal beliefs” make “I expect...” materially different? I was only abbreviating an already too long comment. I’d have said it was emphasis that didn’t change the meaning, rather than a qualifier. But feel free to re-insert the qualifiers, and I’ll still stand by it.

I do not think I think it means what I think you think I think it means.

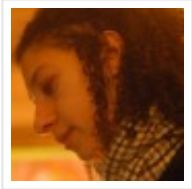
November 28, 2010 at 12:32 pm



She-mathematician

Just for the record, Jennifer Ouellette isn't a physicist (although she hangs out with that crowd). Can we do a little better next time at finding pictures of female physicists?

November 28, 2010 at 11:23 pm



Chanda

A few of us have been looking to get in touch with Reva Williams for years now. If anyone knows how to (or Reva if you are reading this) please get in touch with me.

Great post. 😊

November 29, 2010 at 1:08 am



becca

Nullius in Verba- Including the “according to my own personal beliefs...” increases the scope for interpreting the question specifically as one deems appropriate.

Personally, I would answer the question “I expect...” based on what I thought was most relevant to why they were answering the question. If I was asked the question using that phrasing when I was taking a particular physics course, I would assume they meant physics performance in that specific context (as you also assume), so I would score as “definitely agree”, using this interpretation:

* “I expect men to generally do better in this physics course than women” (this, as you point out, would be an evaluation based on facts, not opinion; in this sense, I expect women to do worse in physics than men)

On the other hand, if the question were prefaced with “according to my personal beliefs...”, I would feel more comfortable defining the question in a specific way that seems most salient to me. Such as:

***I do **not** expect men that become physicists to generally do better at physics than woman that become physicists” (I believe that, on average, women who make it through a lot of small biases against them during the educational process, come out more competent than men. In this sense, I expect women to do better in physics than men)

Given that the question did allow for either way of interpreting the question, I find it Very Interesting that you, personally, can ONLY see the former interpretation. The interpretation under which the question becomes a matter of simple fact. That happens to confirm the predominant cultural stereotype.

(sidenote: These subtle differences in question format are NOT viewed as minor things in the field of psychology, for good reason. Even if they wouldn't affect how *you* view the question, if adjusting the question affects how *anyone* views it,

it's a relevant variable. What you have done is the equivalent of examining a drug response at 3 hours instead of 3 days. You might easily get completely different results from the two different questions.

Psychology is generally only cargo-cult science when you *ignore* important, but apparently subtle aspects of experimental interpretation.)

November 29, 2010 at 12:30 pm



jdmimic

For those that asked about the men scoring worse in the affirmation group than the control, they discuss it much more in depth in the supporting online information than they do in the paper itself. Sadly, they do not provide all the raw data, so the ability to exam their data is limited.

@Becca: I interpreted that question the same way Nullius did. It is not a factor of cultural stereotype, but of an ambiguous question as you pointed out yourself that there is more than one way to interpret it. When I read it, I found myself unable to adequately answer it for myself because they did not adequately define parameters. I admit though, that I have been considered borderline autistic when it comes to social interpretations. The "according to your own beliefs," portion of the question simply confuses me as my own personal beliefs have nothing to do with who will do better in a subject, nor does it address the reasons for said performance.

If the point of the question was to determine what the person thought of women's ability in physics compared to men, why did they not just ask that? I would think that asking the question as, "Do you think women are as capable as men in physics?" would be much more unambiguous than, "According to your own personal beliefs..."

I personally detest taking psychological exams because every single one I have taken had almost every question so ambiguous and required such simplistic answers that I could not fairly answer the question and I felt I was lying. I have taken tests before more than once and gotten very different results simply by choosing a different interpretation of the questions the second time around. Both times I answered correctly as best I could given the interpretation of the question.

Since you and Katt seem to have some experience, or at least knowledge in this regard, perhaps you can enlighten me as to why psychological testing seems so rampantly unclear.

And just to clarify: I personally believe that there is little difference in men and women in physics capability, but a large difference in social pressures and training that results in a different result. How should I then answer the question? Not trying to be argumentative, just curious as to your answer because I would not know how to answer it.

November 29, 2010 at 2:14 pm



Bram

I'm mainly interested in the used statistics and the problems with it. Let's for instance assume the idea of stereotype threat to be real: do the figures and used statistics, as criticized by Nullius in Verba, give way to mentioned conclusion, i.e. '15 min of writing on important values will close the gender gap?'

Nullius, you said:

"For example: there are fewer women in senior positions, but this seems mainly to be due to their unwillingness to play office politics or take on all the bureaucracy – a characteristic shared by a lot of men who also do not 'get on'."

Do you have data to back that up?

November 29, 2010 at 2:39 pm



Nullius in Verba

#49,

That is very interesting. And you're right, I can only see the former interpretation. In fact, even after reading your explanation, I still don't see how the phrase "according to my own personal beliefs, ..." turned out to be equivalent to inserting "...that become physicists..." into those two places. I would have expected a qualifier that spoke of "my own personal beliefs" to have been contrasted with somebody else's beliefs, their non-personal beliefs (whatever those might be), or something other than a belief. (Deduction? Hypothesis?) I'm also interested to see that you interpreted "I expect men to generally do better in physics than women" without the qualifier to be specific to taking a particular physics course. I can sort of see how you do that one – the word "physics" is interpreted as the name of the course rather than the name of the science – but I'm afraid I'm still mystified as why you prefer one to the other.

Your wider interpretation ("I believe that, on average, women who make it through a lot of small biases against them during the educational process, come out more competent than men") is interesting for another reason. It's something like saying the top 15% of women (who make it to be physicists) are better on average than the top 50% of men (who make it to be physicists). (Don't take the percentages too literally.) It's obviously almost certain to be a true statement. But it doesn't say what I think you would *like* it to say – that women generally are at least as capable as men.

I'm also not sure why "do better in physics" got changed to "are better at physics". If women are subject to many small biases against them (something I'm not convinced is still generally true, but let's pass on that) then I don't see why this should stop when they graduate. They would do no better in physics after graduation (when they're even more outnumbered than they were as a student) than they did before, on a physics course. Professionals have less protection against academic politics than students do. Also, I'd have said that dropping out of physics early on doesn't really count as "doing better".

I agree that the question wasn't worded very precisely, and as with most language, it can be parsed multiple ways.

I am genuinely interested in the way you managed to re-interpret the statement along such completely different lines, so please don't take my mildly humorous approach the wrong way. It has given me some food for thought. But I'm also genuinely puzzled at your interpretation. It seems to me that it is a bit like taking the statement "according to my personal beliefs, I expect patients given a low dose of strychnine to make better recoveries than those given sterile water" and using the first clause as an opportunity to redefine "recovery" to mean those patients that *survived* the strychnine turning out to be stronger and healthier in the long run than those survivors dosed only on water. An interesting and lucrative future in medical research awaits, I'm sure!

#51,

The 'data' is anecdotal, I'm afraid! The people I see getting promoted are the ones that are known for playing politics. I did say it was from my own personal experience.

November 29, 2010 at 3:30 pm



Jatila

Great article, and I will try this with my students! I offer one correction, for the record: Jennifer Oullette is NOT a physicist; she is a science WRITER. A physicist is someone who has a PhD in physics. Jennifer is an excellent writer, but she does not have a PhD in physics. (Being married to one does not substitute for having gone through the initiation rite of getting a PhD in physics.)

November 30, 2010 at 1:50 pm



Cam

Why are they dressing up one of the typical self esteem excersizes that have been used in schools since the 1970's and pretending that it is something new?

November 30, 2010 at 1:54 pm



tedda

#52 Nullius in Verba, and #49, Becca:

I also felt that including "according to my own personal beliefs..." completely changed the question. Without it, the question seems simple. I DO expect men to generally do better in physics than women, because I'm aware that statistically, men are likely to outperform women in science courses/careers/etc. You'd have to be living under a rock to put any other response, because the disparity is fairly well documented. However, including "according to my own personal beliefs..." makes the question more ambiguous – though it's a subtle difference, I now have trouble interpreting the question. This detail makes me inclined to consider the fact that MY belief is that women are just as *capable* as men in the field of physics, whether or not the actual outcome is that men "do better in physics" than women. I'd be more inclined to put my response somewhere in the middle in this case. That minor detail means that the question does not necessarily require that you respond in terms of "according to what we actually see..."

...I feel that the question would mean more, and the difference between including or excluding “according to my own personal beliefs...” would be more notable if it said “...are better at...” rather than “...do better in...”, as the latter implies that you should comment on what is actually seen rather than what might be seen if there was no bias, or what innate ability women and men have.

Anyways, this discussion has been great, and is motivating me to think critically in future scientific endeavors (I’m still a student).

November 30, 2010 at 4:52 pm



Zaotar

Very interesting comments Nullius in Verba. I was writing about similar issues on the Slate page:

<http://www.slate.com/id/2276066/>

But I see you’ve gone more in depth. As a lawyer who works on a lot of litigations involving consumer survey evidence, social scientific methodology is a very interesting subject for me. In my experience, social scientists are often either incompetent with statistics or all-too-willing to distort their methodology to support conclusions they want to reach. Rarely do the authors have much skill with proper statistical methodology, something that is very difficult in the social sciences even when you genuinely intend to do it right. Thus it’s important to be critical of social science studies.

Generally the control structure is where all the action is, when you are looking at the validity of study data. I actually like the control structure they used, at least in theory. However, I think there are also real sample size problems — as the results are reported, at least, the original study is surely much more careful. You noted, as I had, that you are talking about control/experimental groups of 55 to 60 women, roughly. But also notable is the fact that the grade differences were only observed among women who got Bs and Cs. So that cuts the groups down even further. I suspect we are talking about comparing groups of 30-40 women, when you are looking at the group the study claimed to find a significant disparity in. But really — comparing a 30 to 40 sample size? That’s really pushing it. Just a few women could knock the number substantially one way or the other, based on random chance alone.

The other issue that you alluded to, but which should be pointed out, is that there were a total of 399 survey participants, 116 being women. Pretty plainly the study took ALL women, and then filled the rest up with men. Otherwise they would have chosen a 100/100/100/100 structure, which obviously is what they had ideally wanted (these studies being generally conducted in groups of 100). But there were probably not enough female students to get 200, so they had to scrape the barrel and then fill up the rest. The problem with doing that is that you are distorting your comparative samples.

Finally, the bizarre reduction in the men’s scores on the exam, within the experimental group, needs to be addressed. When men wrote a short essay affirming their own values, it actually LOWERED the men’s scores on physics

exams, relative to writing a short essay on other peoples' values? REALLY?! Incredibly unlikely. The entire purpose of a control group is to indicate whether your data reflects alternative explanations to causation, such as flaws in your methodology. Getting such a peculiar result in your data indicates that either (a) something in your study was off-kilter; or (b) there is a really weird, UFO-like causation issue that you have genuinely stumbled upon here. Almost certainly the former, when it comes to the reduction in test scores. But those reporting the research in popular media just gloss over this bizarre data (evidently the study itself goes into greater detail on it, which is what I would expect). It's a flag that something is out of whack with their survey data, even when (as with a complete experimental group of male subjects) the authors were using a vastly larger comparative sample size (here, about 140 men in each group). Suggests something truly goofy was afoot here.

Basic summary of my thoughts, to riff on Hume: Extraordinary claims require extraordinary evidence to confirm them, and there seem to be a lot of problems with the evidence here.

November 30, 2010 at 6:45 pm



becca

Nullius in Verba-

I was imagining myself answering the question in the context of being given the question during a physics course, as the subjects were. In *that* context, I think I would have been likely to interpret 'physics' as 'this physics course'.

I absolutely see what you are saying about the medical treatment- it's a good analogy. While my interpretation of the question might be wonky, the clause "according to my personal beliefs" kind of *encourages* one to use a wonky interpretation if one so desires. And, given what tedda says, my basic point- that the nitty-gritty of how you ask the question matters- is probably plausible.

Zaotar- as a scientist, reading armchair philosophers oh, I'm sorry, lawyers comments about scientific studies always interests me. Rarely do the lawyers have much skill with properly reading journal articles, such as we teach in the first year of most PhD programs in journal clubs.

For those that might not have access to the article, on the topic of men's scores, the authors have this to say:

"Unexpectedly, affirmation negatively affected men's exam scores, but, unlike the positive effect for women, this effect was not predicted, was not replicated for the end-of-semester FMCE score, and did not change men's letter grade distribution. In contrast, the affirmation's positive effect on women was significant for all outcome variables (SOM text), suggesting that the reduced gender gap observed in this study is based more robustly on the affirmation's positive impact on women than on its negative impact on men."

Now this does not *explain* the negative men scores. But not necessarily indicative of anything off-kilter or UFO-esque.

One aspect of this study I think is worth considering is how the exam scores differed from the FMCE scores. An important variable in the two parameters is

that students were explicitly told the FMCE scores would *not* affect their course grade. That makes one aspect of the findings of the researchers, even more intriguing:

“After controlling for prior background (prior SAT/ACT Math or beginning-of-semester FMCE scores), the affirmation closed the “residual” gender gap on in-class exam scores by approximately 61% and entirely eliminated the gap on the FCME.”

In other words, when the students knew their performance was important to their course grade, the women benefited *less* from the affirmation than when the students knew their performance was not important (either that, or there is a substantial difference in the physics abilities being tested by the exams vs. the FCME that is confounding things). In addition, the male students were *hurt* by the affirmation exercise for the exams (where their performance mattered) but *not* for the FCME (where their performance didn't). I wonder if test-anxiety led to self-defeating behavior in both groups? I haven't read any of the relevant psychological literature, but I suspect there's some addressing this pretty rigorously (test anxiety is a pretty replicable phenomenon, at least anecdotally).

All that said, I am inclined to agree with the commenter that pointed out the *specific* personal values that were chosen may be playing important roles in priming behavior. That, of course, does not in any way imply that the intervention isn't *useful* for what the authors say it is, just that we can't know how it's working exactly in any individual's mind (though it seems likely to me, given the critical variable of endorsement-of-stereotype, at least in *some* people this may be working as the authors suggest. Psychology is messy that way- if you saw the exact same effect in every individual you wouldn't need statistics).

December 1, 2010 at 12:04 pm



Zaotar

I agree that lawyers, generally speaking, rarely have much skill in dealing with scientific literature. But at least in my field of expertise — patent and intellectual property litigation — that certainly ain't so. I've seen scientist after scientist taken apart by lawyers who knew the science, and the facts, much better than the 'expert' did. That's particularly common when scientists attempt to opine outside of their narrow field of technical specialization, something that is not uncommon (and certainly rather common when it comes to the social sciences — for matters that are heavily politicized, in fact, it's basically a default assumption that the statistics are going to be soft).

At any rate, I certainly don't want to initiate a social scientist v. lawyer battle (there is no profession I respect more than that of the scientist, albeit moreso the 'hard' sciences — I'll take a scientist over a lawyer any time), but thanks for posting the authors' comments about the decline in test scores. With respect to their first comment, however, that “this effect was not predicted,” well that's precisely what you expect of comparison data that indicates an apparent methodological problem. The unusual effect is not supposed to be predicted. The fact that it shows up, contrary to what anybody would reasonably expect could be true based on the methodology and sample size, is your indication that something is potentially amiss in the methodology.

The other comments you quote on this issue are well-taken, but a bit limited in scope. After all, that “exam score” graph appears to be a compilation of all four exam scores, and for an approximately 140 sample size (total men in the “affirmation” group) to show such a significant decline across *four aggregated exams* — that is really remarkable and implausible. Pointing out that the ultimate grade distribution didn’t change isn’t an overly significant rejoinder to the fact that their exam performance (which the grade is derived from) noticeably suffered, something that is hard to attribute to anything other than methodological problems.

Now, as I say, given that the men’s sample is far larger than the women’s, it was rather surprising that such a variance could be generated at all in the men’s data — it was a pretty significant drop in aggregate exam scores, based on a correlation (affirmation of values = decline in exam scores) that almost nobody thinks could truly be a causal result. If that kind of presumably non-causal change in test results can be generated for the men, even with their much larger sample size, it’s not exactly surprising that a significantly smaller sample size would also generate peculiar variances. It’s also not surprising that, if the variation in performance was stacked towards one smaller group rather than the other (i.e., higher performers in one group), that same variation would be replicated in the other group. That’s precisely what you would expect. Now it’s true that the mens’ FCME scores didn’t decline much (they appear to have also declined very slightly in the affirmation group, if that bar graph is any indication, just not nearly as much as the exams), so there appears to be unequal correlation of FCME scores and exams for men. Which, well, who knows what the hell that means. The data is rather baffling on several fronts (why did the benefit just hit B and C grade women, for example? Why not the numbers for Ds and As? You could formulate any one of a thousand hypotheses, in the absence of data, but that is also an “unexpected” result that should be sending up methodological flags).

My overall feeling is that this study seems promising enough, relative to the minor cost of just writing an essay, that it surely warrants more detailed and careful follow-up studies, with rigorously defined controls and larger groups for comparison. There’s no doubt that, if true, writing a cheezeball value essay increases grades so much, then it should be standard-practice in many different educational contexts. Why the heck not. But I’m highly skeptical, unless the data can be convincingly replicated with more typical sample sizes (at least 100, and preferably 200, is what I’d hope for).

Incidentally, as you have access to the study itself, could you post the subgroup samples that are being compared in generating the comparison data? Control/experimental for male/female, and for the grades, the B/C grade women control versus B/C women grade experimental? I’m guessing it breaks down 55/60 women in each group, and 30/40 women in each sub group within the B/C range, but I have now way to tell for sure from the stuff available on the web.

December 1, 2010 at 4:05 pm



becca

Zaotar- I don't care how well lawyers think they can take apart anyone, you can't be very good at interpreting science without reading the actual paper. We ain't talking rocket surgery here 😊

The authors note: "this negative affirmation effect on men's exam scores was not significant when the analysis was conducted with the beginning-of-semester FMCE scores (instead of SAT/ACT Math scores) as the covariate"

That is, it appears that they had, presumably through random chance, male students with poorer initial conceptual understanding of physics (assuming the FCME measures what it claims to) in the affirmation group than the control group.

But I think it's interesting to think about how affirmation could actually decrease scores.

Here are the values that were actually part of the exercise: "being good at art; creativity; relationships with family and friends; government or politics; independence; learning and gaining knowledge; athletic ability; belonging to a social group (such as your community, racial group, or school club); music; career; spiritual or religious values; and sense of humor".

Now, hypothesize with me here. My friend who is TAing biology classes reported to me that ~1/3 of his female students have come to office hours, and none of his male students. My immediate, blatantly biased response was 'that's because your female students are working harder'. But then I started thinking about why else that might be. Is it possible that female students are rewarded more for appearing to work hard? Or, conversely, male students are punished more for seeking help? I don't mean punished directly in grades, or anything, just more or less socialization pressure.

Now, that might not seem relevant. But I'd be rather surprised if there *weren't* some gender differences in the particular values each group decided to write about.

Is it possible that either male students were in the affirmation condition were *more* likely to write about independence, or simply that males writing about independence were more likely to avoid going to office hours than the female students, leading to worse exam performance?

Would this constitute something wonky with the study, or an additional variable of interest well worth studying (i.e. what list of values for affirmation has the best increase in *everyone's* scores)?

Of course, it's probably simpler (albeit more stereotyped) to hypothesize that females were more likely to choose to write about learning and gaining knowledge and that anyone that wrote about that particular value had higher increases in performance.

If we knew what they wrote about, we'd know what hypotheses to test.

A *GOOD* study, the kind that generates press and gets published in Science, *should* be the beginning of new investigations, not the last word. So I think you're ideas about unexpected results being *automatic methodological red flags* are horribly naive and flawed. If, on the other hand, you have some specific

qualms about the statistical analysis that you can explain, that would be very interesting and perhaps edifying to myself and others.

Of course, as the authors note, if you want to use these data and you're really worried about the men- just give them the control condition.

Here's what I can find regarding the actual numbers:

The analysis is based on 283 men and 116 women

In affirmation condition: 178(M) 69(F)

The FCME data were based on 137(M) and 55(F) in the affirmation condition and 75(M) and 41(F) in the control.

For the grades- you might not see the effect in D students simply because there are so few. But why it wouldn't boost some students to As, I can't fathom- I agree that is weird.

Here are the specific percentiles: "A large majority of women in the control condition (55.8%) earned a grade in the C range (including C-, C, and C+), with only 23.1% earning a grade in the B range (including B-, B, and B+). The percentage of Cs was reduced to 40.8% among women in the affirmation condition, and Bs increased to 36.8%. This difference in the percentage of women getting Bs and Cs across the two conditions was statistically significant [$\chi^2(1, N=91)=4.07, P=0.04$]. There was no difference in the distribution of Bs and Cs for men as a function of affirmation condition [$\chi^2(1, N=202)=0.02, P=0.88$]."

December 2, 2010 at 1:10 am



Hope

@ Jatila (#53): You are correct that Oullette is not a physicist; she's a writer. The fact that she lacks a PhD in physics is not the reason that she's not a physicist, however. You can be a physicist without having a PhD in the field. And for the record, her husband doesn't have a PhD in physics either. 😊

December 2, 2010 at 2:18 am



Zaotar

Becca — thanks for answering my information request, which is very helpful. I don't know to unpack their statistical significance calculations — they just give the numbers, not their calculations — but a .04 number is generally only slightly on the side of statistical significance (my understanding is that usually .05 is the cutoff between significant/insignificant, with less than .05 being significant).

So if their math is correct, then it appears there is most likely some statistical significance to the observed variation, but that is (I would say) a far cry from asserting that the entire observed variation therefore has been shown to be caused by the test condition. If you are looking at a total sample of 91 broken into two groups (approximately 55 and 36), that's a pretty small sample, which should be approached with caution, especially if the much larger groups are also showing some whacky data (the authors give the B/C distribution probability for the men,

but they don't calculate the statistical significance of the underlying grade decline, which should have been easy enough, and is the real issue).

By sending up "methodological red flags," I don't mean to say that the data is junk, just that one should be careful and skeptical.

Again — thanks for taking the time to help me out with the underlying data here. Nice to hear your thoughts.

December 2, 2010 at 4:41 pm



Rebecca

So who plagiarised who here? Did you plagiarise Cordelia Fine, who has almost the same word for word outline about this in her latest book "Delusions of Gender" or did she plagiarise you?

December 4, 2010 at 5:45 am



Ed Yong

Cordelia Fine's book is on my Christmas list but I haven't read it yet. But don't take my word for it – her book was published on 30th August 2010, and presumably written well before that. The research that I discuss in this post was published three months later on 25th November 2010. For plagiarism to have occurred, Fine would have to be capable of either time travel or clairvoyance and while she's undoubtedly a great writer, I doubt she has either of those skills.

Of course, all of this information is readily available. I cited the research I wrote about. Google/Wikipedia reveal the details of Fine's book. Given it would take less than a minute to establish both facts, the only reason I can think of for failing to do so before making the pretty serious accusation of plagiarism, would be if someone was a bit of a wanker.

Oh.

December 4, 2010 at 11:22 am



Djoke Veldhuis

Ed

Abi Millar from City University's Science Journalism team did a nice review of Cordelia's book a few weeks ago — <http://www.elements-science.co.uk/2010/11/delusions-of-gender-by-cordelia-fine/> — that should whet your appetite 😊

Regarding plagiarism, now I am reporting from science events, lectures etc. it never ceases to astonish me how I can read a review / blog weeks after I've written an article on subject 'x' and find that I had put in the same quotes by the same people who spoke on a particular subject. I suppose to some extent this is inevitable — a good journalist should be able to pick out the most salient points / theories / results / implications of a study or a lecture.

Conclusion: it's like you say based on the sources you're using there are only so many different combinations you could come up with, I would pay no heed to unfounded accusations of plagiarism



December 4, 2010 at 11:47 am



Robbie Lamons

Nullius in Verba wrote some lovely lines that I would like to be able to quote with a real name, and if not, to sew into samplers, carve into stone, or engrave onto brass plaques to hang in schools everywhere with Nullius in Verba as the author.

35.

“And one of those biases is confirmation bias – we examine theories we don't like or that conflict with our current beliefs more critically than those that we are already inclined to believe. In science, we must guard carefully against it, and the first and most important step to doing so is to recognise and acknowledge that we have it. More, that all the other scientists have it, and that despite all our efforts to keep it out it inevitably pervades even published and widely-accepted results.”

January 7, 2011 at 5:05 pm



Ann Onymous

I can see that this would work. I (female) studied physics in a class with Male to Female ratio of around 15:1. The most difficult thing about this was not the subject, but the perception that male class members seemed to know what they were doing, notably during lab sessions.

The males in the class would be very confident and sure of themselves in approaching practical work, whereas generally the females were tentative and felt that they would mess up or break something. Usually the males would take over the task and the female members would be pushed to the sidelines to watch. When the practical work was over, this left a feeling of not having participated enough to have fully understood the task, which lingered over the rest of the course. This would lower self esteem and turn into an inferiority complex. So it makes sense that a confidence boosting exercise would have this effect on women.

Something else which has been shown to close the gender gap, in stereotypically male subjects, is to have single sex classes.

January 16, 2011 at 12:27 pm



Nickolas

Holy Crap!!! I'm going to the University of Colorado at Boulder right now and I had to take this exact same survey! I was wondering why my friend told me that he had to write about why values were important to other people compared to me writing about my values. I'm ruining the results by reading this!!!

OMG!@#\$^@!#\$\$111!

January 31, 2011 at 5:38 pm



Ed Yong

Ha! Not Exactly Rocket Science: inadvertently ruining experiments since 2006.

January 31, 2011 at 5:43 pm



Katwoman

@Ed – I wonder if the timing and/or number of writing tasks play a role in the results? For example should the lapsed time between writing tasks not go beyond 2 weeks? Should the first writing task be given with the first month? You get my drift.

May 12, 2011 at 7:47 am



Elly

This is an interesting concept and article, however, as a woman in science, and most specifically in physics, I have some points for discussion.

Firstly, there was only one study conducted alone, which does not equate to success or failure in any scientific endeavor. Furthermore the “random sampling” could have by chance had males which were better prepared from high school (this is an introductory course, so AP physics or quality of high school physics classes has a HUGE impact) paired with women who were not as prepared from high school lumped into the same category and vice versa. This is apparent in the graphs provided. The male statistics were ignored. If you look at the male control group compared with the experimental group, you can actually see the scores go down, not up, with the value affirmation. More studies need to be done before this is valid. I do hold that it is important to be confident in yourself, but I am not convinced that this exercise would have that profound an impact.

Secondly, on a personal note, when has the concept of “equality” among men and women, meant special treatment for one group. I find that increasingly so, women are given special treatment to make it “fair” and “equal”; specifically in this case, this article implies that women were given extra study sessions etc, to help fill the gap in the past. I find that this preferential treatment does the exact opposite, psychologically, of what is aimed to be accomplished by the writing task. The extra treatment isolates that whole group of the class, instead of promoting collaboration, impartiality and fairness in scholarly pursuits. By providing extra help specifically to women, it REAFFIRMS notions that they need more help, are not as naturally talented, and require extra attention in order to be “on par” with the men.

September 8, 2012 at 2:55 pm

To all the male haters posting here: the point of this study doesn’t appear to be, “Now we can say beyond a shadow of a doubt, for all time, that this study definitively proves that there’s bias against women in the sciences.” Most science



m23

doesn't stand up to that sort of thinking (re-read your Kuhn if you're unclear on the concept).

The point of the study is, "Working in a well established realm of study, stereotype threat, this exercise helped women perform better in an extremely male-dominated science education environment." Someday, someone can come up with some sort of dark matterish theory to explain why. For the immediate now? Increasing performance among minority students (and in science, women are a minority) is a good thing.

If you honestly believe there are no societal factors contributing to the dominance of men in the sciences, art, etcetera, then wow. Start at the beginning, do your homework, and THEN jump into this conversation.

September 26, 2012 at 7:03 pm

